CRM 88-77 / May 1988

AD-A199 868

RESEARCH MEMORANDUM

ASSESSMENT OF DON SHORE BASE READINESS ANALYSIS

Martha E. Shiells
Richard C. Tepel

DTIC
SELECTED
OCT 2 5 1988
D
D

A Division of **CNA** *Hudson Institute*

CENTER FOR NAVAL ANALYSES

88 1025 006

Approved for public release;
distribution unlimited

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRM 88-77	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Center for Naval Analyses	6b. OFFICE SYMBOL (If applicable) CNA	7a. NAME OF MONITORING ORGANIZATION Office of the Chief of Naval Operations
6c. ADDRESS (City, State, and ZIP Code) 4401 Ford Avenue Alexandria, Virginia 22302-0268		7b. ADDRESS (City, State, and ZIP Code) Navy Department Washington, D.C. 20350-2000

8a. NAME OF FUNDING / ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-87-C-0001

8c. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 65154N	PROJECT NO. R0148	TASK NO.	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

Assessment of DON Shore Base Readiness Analysis

12. PERSONAL AUTHOR(S)
Martha E. Shiells, Richard C. Tepel

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Year, Month, Day) May 1988	15. PAGE COUNT 54

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Bases (structures), Facilities, Mathematical models, Mathematical
12	03		predictions, Naval shore facilities, Readiness, Regression analysis,
15	05		Statistical analysis .

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research memorandum documents the Center for Naval Analyses' assessment of the Price Waterhouse shore base facility condition readiness model. The accuracy and and reasonableness of the model's predictions are assessed. Suggestions are made fo revising the presentation of the model's results and for refining the model.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.

All other editions are obsolete.   SECURITY CLASSIFICATION OF THIS PAGE

9 June 1988

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Memorandum 88-77

Encl: (1) Center for Naval Analyses Research Memorandum 88-77,
              Assessment of DON Shore Base Readiness Analysis, by
              Martha E. Shiells and Richard C. Tepel, May 1988

1.   Enclosure (1) is forwarded as a matter of possible interest.

2.   This research memorandum documents the Center for Naval Analyses'
assessment of the Price Waterhouse shore base facility condition
readiness model. The accuracy and reasonableness of the model's
predictions are assessed. Suggestions are made for revising the
presentation and for refining the model.

LEWIS R. CABE
Director
Program Analysis Group

Distribution List:
NAVWARCOL
NAVFITWEPSCOL
OP-44
OP-442 (10 copies)

Accesion For

NTIS CRA&I	☑
DTIC TAB	[ ]
Unannounced	[ ]
Justification	

By

Distribution

Availability Codes

Dist

A-1

ASSESSMENT OF DON SHORE BASE READINESS ANALYSIS

Martha E. Shiells
Richard C. Tepel

Navy-Marine Corps Planning and Manpower Division

ABSTRACT

This research memorandum documents
the Center for Naval Analyses' assess-
ment of the Price Waterhouse shore base
facility condition readiness model. The
accuracy and reasonableness of the
model's predictions are assessed. Sug-
gestions are made for revising the pres-
entation of the model's results and for
refining the model.

EXECUTIVE SUMMARY

Price Waterhouse prepared a statistical model that relates Mainten-
ance and Repair of Real Property (MRRP) funding to facility condition
readiness.[1] CNA was asked by the Director, Shore Activities Division
(OP-44) to conduct an independent assessment of this model and to iden-
tify promising alternative modeling approaches.

Developing a resources-to-readiness model for shore base facility
condition readiness is inherently difficult. The Shore Base Reporting
System (BASEREP) data are limited in both quantity and quality. Given
the quality of the data, it may not be possible to find a relationship
between funding and readiness that is statistically significant and
conforms to common sense, regardless of what estimation techniques are
used.

Price Waterhouse should be commended for preparing a model that is
simple and well documented. The CNA study team, however, found signifi-
cant shortcomings with both the statistical techniques used and the
manner in which results were presented and interpreted. These criti-
cisms seem more important because the predictions generated by the Price
Waterhouse model fail tests of reasonableness and of statistical impor-
tance. At the very least, the report should be revised so that the
statistical results are presented with the proper qualifications. The
study team further recommends that a more defensible model be produced--
one that would yield more reasonable predictions.

VALIDITY OF PREDICTIONS

Even without considering the validity of the statistical techniques
used, the Price Waterhouse model can be faulted because it does not make
reasonable and accurate predictions. The model's predictions of facili-
ty condition readiness for all Navy facilities are pessimistic. Sup-
pose, for example, that the level of Replacement and Modernization
Military Construction (R/M MILCON) funding is held constant from 1988
through 1994, and MRRP funding levels are taken from the President's
1988 budget submission. In this case, the Price Waterhouse model pre-
dicts that the percentage of C1 or C2 ratings will fall from 75 percent
in 1983 to 46 percent in 1994. Specialists in base readiness from OP-44
do not believe that this large a decline in readiness is likely.

An examination of out-of-sample predictions demonstrates that the
model's level of accuracy is low at the sponsor/claimant level. These
predictions were made by deleting data for one year, reestimating the

---

1. Price Waterhouse, Contract No. N00600-86-D-3869, Delivery Order 2,
*Development of an Analytical Model Relating MRRP Resources to Facility
Condition Readiness*, Aug 1987.

model, and then determining predictions for the deleted year. The predictions for the deleted year were then compared to actual readiness for that year. In one test, the predicted direction of change in readiness was compared to the actual direction of change. The direction of change was predicted correctly only 42 percent of the time. An alternative test was to consider how well the model performs when compared to a simpler model. The simplest model would predict no change in readiness, whatever the level of funding. Statistics were calculated based on the differences between predicted and actual levels of readiness. These statistics show that the simple model virtually always performed better than the Price Waterhouse model.

SUGGESTED REVISIONS OF PRESENTATION

Although the presentation of the Price Waterhouse model is generally clear and complete, there are a few instances in which the interpretation of results may be misleading. The executive summary states that a 23-percent increase in funding would be necessary to maintain readiness at constant levels. This prediction is based on the individual intercept terms, over half of which are not significant. Also, the predicted break-even level of funding is outside of observed funding levels for over half of the sponsor/claimants. This result should not be presented as if it were fact, with no reference to its statistical validity or to the difficulties with the underlying data.

The final Price Waterhouse model was the result of a lengthy specification search. That is, many preliminary regressions were run and the results of these regressions were used to choose a final model. This is a widely used technique, and some amount of specification search is virtually unavoidable. The statistics associated with a model arrived at by this method, however, must be used with caution. The final model may say as much about the criteria used in the search as it does about the true relationship between funding and readiness. For this reason, the statistics associated with the final model should be reported witn the proper qualifications and used with caution.

SUGGESTED REFINEMENTS OF THE MODEL ,

In addition to revising the presentation of the model, it is the opinion of the CNA study team that the existing model could be improved. A few of the most important problems are discussed here; other problems are mentioned in the outline at the end of the Executive Summary and are discussed in the text. A detailed presentation of an alternative model that incorporates these suggestions is given in appendix B.

The purpose of the Price Waterhouse modeling effort is to relate funding to facility condition readiness. For the model to be successful, it must be based on a meaningful measure of readiness. Whether a measure of readiness is meaningful must be decided by the users of the model. This issue cannot be decided by a statistical test as Price

Waterhouse attempts to do. The readiness measure used in the Price Waterhouse model is the percentage of ratings that are either C1 or C2. This measure gives equal weight to port operations in Long Beach and fire protection in a Reserve Center. Alternative readiness measures would attempt to weight the ratings by measures of the size and importance of the mission and activity. Any weighting scheme introduces complications. If a model that simply predicts the percentage of C1 and C2 ratings provides the information that the Navy needs to allocate MRRP funds, then the complications of weighted readiness measures can be avoided. If, however, a more sophisticated measure of readiness is required, then these complications must be addressed.

The small size of the data sample influences how many variables can be included in the model and also creates pressure to increase the number of years of data used. Further testing needs to be done to determine the best level of aggregation at which to estimate the model: sponsor/claimant, only sponsor or only claimant, or Navywide. If data from different years are consistent, then the accuracy of the model can be improved by using more years of data. Appendix C describes tests of whether different years of data can be combined.

Although the Price Waterhouse model in general has the virtue of simplicity, its treatment of time-series/cross-section effects is unnecessarily complicated. The model combines individual sponsor/ claimant intercepts with variables transformed into differences from lagged values and ratios to average values. As a result, the model is a hybrid of three standard time-series/cross-section models. It is recommended that one or another of the standard time-series/cross-section models be adopted. Combining the different models may produce unexpected statistical results. Furthermore, the coefficients in their hybrid model are difficult to interpret.

The discussion of level and change models in the Price Waterhouse report is misleading. It must be determined whether readiness in one period tends to decline to some fraction of readiness in the previous period. If there is no such decline, then it is correct to estimate the model that Price Waterhouse proposes, with the change in readiness on the left-hand side. If there is depreciation, then the previous period's readiness belongs as an independent variable on the right-hand side of the equation.

The Price Waterhouse model uses a linear functional form. The justification used for this is that even if the true readiness function is not linear, a linear relationship can approximate a curve within a limited region. Although this is true, the results of the estimation are not used to predict readiness changes within a limited region. For this reason, and because there are theoretical reasons to believe that the relationship between funding and readiness is not linear, the CNA study team believes that a nonlinear functional form should be used.

This point is important because the linear functional form may cause the model's overly pessimistic predictions.

CONCLUSIONS AND RECOMMENDATIONS

The major conclusions and recommendations of this assessment are as follows:

- The usefulness of the existing Price Waterhouse model is limited because of the quality of its predictions. Predictions of Navywide shore facility condition readiness are pessimistic; predictions at the sponsor/claimant level are inaccurate. The model should not be used to reallocate funds among sponsor/claimants.

- The following revisions are suggested in the presentation of the model:

  -- The executive summary should be expanded so that it reflects the uncertainty regarding the statistical results.

  -- Coefficients should be interpreted in terms of changes from average funding levels rather than break-even funding levels.

  -- The report should document the model that is delivered to the Navy in a LOTUS spreadsheet.

  -- The statistical results should be qualified even more heavily because the final model is the result of a specification search.

- The following refinements are suggested in the handling of the data and the estimation method:

  -- The Navy should decide what measure of readiness should be used in evaluating the allocation of MRRP funds.

  -- Variables that measure changes in commanding officers and the percentage of leased assets should be added to the model.

  -- Tests should be performed to determine whether data from different years can be combined.

  -- Tests should be performed to determine whether a sponsor/claimant model, Navywide model, or some intermediate model should be used.

-- A more standard time-series/cross-section model should be used.

-- The previous year's readiness should enter on the right-hand, rather then the left-hand, side of the model.

-- Corrections should be made for problems of autocorrelation and measurement error.

-- The relationship between readiness and funding should be nonlinear.

-- A more statistically sound method of weighting the data should be used.

● A model is proposed that incorporates these changes. Appendix B describes this model and shows how it can be estimated.

TABLE OF CONTENTS

INTRODUCTION

Price Waterhouse prepared a statistical model that relates Mainten-
ance and Repair of Real Property (MRRP) funding to facility condition
readiness [1]. CNA was asked by the Director, Shore Activities Division
(OP-44) to conduct an independent assessment of the model and to identi-
fv promising alternative modeling approaches.

Developing a resources-to-readiness model for shore base facility
condition readiness is inherently difficult. (References [2] and [3]
discuss some of the data and modeling problems involved.) The first
Shore Base Reporting System (BASEREP) data were collected in 1982. The
number of bases reporting and the quality of the data have improved each
year. Thus, although data are currently available from 1982 through
1986, using the 1982 and 1983 data is questionable since substantially
fewer bases reported in these years.

Furthermore, the three or four years of usable data are based on
subjective readiness ratings. Base commanders classify readiness at
one of four levels, as listed in table 1. The reliability of these
ratings has been questioned frequently, so much so that a new, more
objective rating system was put into place in 1987 [4]. Finally,
funding must be matched to BASEREP readiness measures at the highly
aggregated spon,or/claimant level. Thus, it is not possible to observe
that the readiness of a specific facility increased when funds were
spent on a particular project to upgrade that facility. This sort of
effect may be swamped by other changes in funding and readiness over an
entire sponsor/claimant. Any modeling effort must be judged with these
difficulties in mind. Given the quality of the data, it may not be
possible to find a relationship between funding and readiness that is

TABLE 1

READINESS RATINGS

Rating	Criteria
C1	The asset has fully met all demands placed upon it in the mission area.
C2	The asset has substantially met all demands of the mission area with only minor difficulty.
C3	The asset has marginally met the demands of the mission area with major difficulty.
C4	The asset has not met vital demands of the mission area.

statistically significant and conforms to common sense, regardless of
what estimation techniques are used.

Given these difficulties, Price Waterhouse has developed a model
relating readiness to funding. They should be commended for documenting
their model clearly and completely. The review process was facilitated
by the high degree of integrity shown in the report and in discussions
with their analysts. Their model has the virtue of simplicity, and
their work in assembling the data base is admirable.

The CNA study team, however, found significant shortcomings with
both the statistical techniques used and the manner in which results
were presented and interpreted. These criticisms seem more important
because the predictions generated by the Price Waterhouse model fail
tests of reasonableness and of statistical importance. At the very
least, the report should be revised so that the statistical results are
presented with the proper qualifications. The team further recommends
that a more defensible model be produced--one that would yield more
reasonable predictions. Given the shortcomings of the data, however,
there can be no assurance that refinements to the model will improve the
results.

The first section that follows examines the forecasts made by the
Price Waterhouse model. It assesses the degree to which the predictions
seem reasonable and the statistical significance of the predictions.
The second section suggests ways in which the presentation of the exist-
ing model could be improved. The third section investigates problems
with the existing model and suggests how they could be corrected. Both
data problems and problems in the estimation methods are discussed. The
first appendix shows how the coefficients in the model can be interpret-
ed. The second appendix contains an alternative modeling approach,
including details on how to implement the suggested estimation proce-
dure. The third appendix discusses possible tests for whether data
generated using the new readiness-rating criteria can be integrated with
the existing data.

VALIDITY OF PREDICTIONS

Two criteria can be used to assess the validity of the model's
predictions. First, one can ask whether the predictions seem reasonable
and whether they fall within the realm of what a knowledgeable observer
might expect. Second, one can look at various statistical tests of the
model's predictive powers. In this section, the Price Waterhouse model
is measured against criteria of both types. The model does not perform
well in any of the tests.

Navywide Readiness Predictions

First, the model's predictions of facility condition readiness for
all Navy facilities are examined. Figure 1 was generated by the Naval

Facilities Engineering Command (NAVFACENGCOM)(Code 1003) using a LOTUS model provided by Price Waterhouse. It should be noted that the LOTUS model is not the same as the model documented in [1] (this will be discussed further in the section on Suggested Revisions of Presentation). The graph depicts the percentage of facility condition readiness ratings that are C1 or C2. Projections are made for four alternative paths of future funding. In each case, the level of Replacement and Modernization Military Construction (R/M MILCON) funding is held constant from 1988 through 1994. Future years' MRRP funding is taken from the Navy's budget request in one case and from the President's budget submission in the second.[1] The remaining cases illustrate real growth of 3 and 5 percent over the Navy's budget request.

Legend:
△ 5-% real growth
◇ 3-% real growth
+ Navy's budget request
□ President's budget submission

X-axis: Fiscal year (83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94)
Y-axis: Percent of readiness (44 to 76)
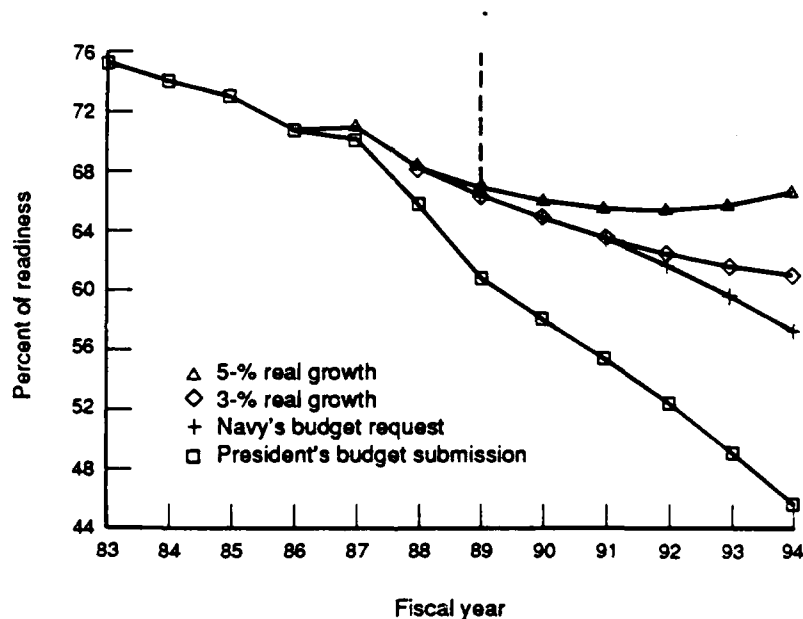
FIG. 1: PROJECTED FACILITY CONDITION READINESS UNDER FOUR ALTERNATIVE FUNDING LEVELS

1. 1993 and 1994 figures were estimated in both cases to show a 1-percent nominal growth of MRRP funding.

The model predicts that even if 5-percent real growth in MRRP funding could be achieved, the percentage of C1 or C2 ratings would decline from 75 percent in 1983 to 67 percent in 1994. Under more realistic funding assumptions, the decline in readiness assumes disastrous proportions. The funding in the President's 1988 budget submission would cause readiness to fall to 46 percent by 1994. If revisions to out-year fundings in the President's budget, Congressional cuts, and Gramm-Rudman cuts were taken into account, the decline in readiness would be even more dramatic.

It is not difficult to derive funding paths that reduce the number of C1 and C2 ratings to zero. For example, if MRRP and R/M MILCON funding were cut in half in FY 1988 and held at that level thereafter, projected readiness in 1994 would be -6.1 percent. The properties of the model that allow it to project negative readiness levels are discussed further in the section on Estimation Problems and in appendix B.

Predictions for Individual Sponsor/Claimants

This section examines the accuracy of the model at the sponsor/claimant level. Although the errors at the sponsor/claimant level may counteract each other to produce a relatively low Navywide error, errors at the sponsor/claimant level are important if the model is to be used to apportion funding among the sponsor/claimants.

As an example, consider the results in exhibit III-1 in [1] (presented in table 2) that lists the estimates of funding required to maintain the present level of readiness. The required funding needed to maintain readiness for sponsor 4/claimant 72 is over twice as high as the average of its previous funding. Conversely, the required funding to maintain readiness for sponsor 5/claimant 61 is only about 63 percent of the average of its previous funding. Both sponsor/claimant pairs have almost the same average previous funding. The estimation results for equation 4 in appendix B in [1] show that the terms that create the difference in required funding have very large variances. Therefore, it cannot be said with much statistical confidence that the required funding is different between the two pairs. Using the model to reallocate funds would lead to a substantial reallocation of resources with very little statistical support that readiness would be improved.

Table 2 examines in more detail the reliability of the predicted break-even funding levels given in exhibit III-1 in [1]. Break-even funding is the funding level at which readiness remains constant from year to year. The readiness model that Price Waterhouse estimates is given by:

$$\Delta R_{S/C,t} = I_{S/C} + M \left(\frac{F_{S/C,t}}{F_{S/C,avg}} \right) , \tag{1}$$

-4-

TABLE 2

VALIDITY OF BREAK-EVEN FUNDING PREDICTIONS[a]

Claimant	Sponsor	Observed range of funding[b]		Predicted break-even funding[c]	Prediction in range?	Intercept significant?[d]
11	1	23.0	39.8	30.9	Y	Y
11	10	17.7	33.7	19.2	Y	N
18	27	65.2	107.0	102.0	Y	Y
23	4	28.1	47.7	39.2	Y	Y
25	4	131.6	136.3	214.5	N	Y
30	2	14.9	16.1	11.9	N	N
60	1	0	1.2	1.1	Y	N
60	2	4.8	26.4	19.6	Y	N
60	3	67.6	120.6	91.5	Y	Y
60	5	80.8	120.0	152.4	N	Y
60	16	2.7	5.6	1.0	N	N
61	2	1.8	2.1	1.7	N	N
61	3	4.1	6.1	5.5	Y	N
61	4	2.1	3.0	4.2	N	N
61	5	14.0	22.3	11.4	N	N
62	1	49.3	99.2	70.0	Y	Y
62	5	46.4	58.7	50.3	Y	Y
62	16	0	3.8	3.8	Y	N
70	2	17.2	22.1	30.9	N	Y
70	3	86.1	93.4	109.9	N	Y
70	4	28.2	40.0	31.5	Y	Y
70	5	97.9	136.8	154.7	N	Y
72	4	11.2	24.0	39.0	N	N
72	5	14.8	27.4	19.7	Y	Y

---

a. All funding amounts are in millions of FY 1988 dollars.
b. Observed funding is calculated from data in [1], appendix A, by
   summing the current year's MRRP and the previous year's MILCON for
   the years 1984 through 1986.
c. Predicted break-even funding is from [1], exhibit III-1, and was
   calculated using equation 2.
d. Statistical significance is tested at the 5-percent confidence level
   using the t-statistics from [1], appendix B, equation 4.

where

$\Delta R_{S/C,t}$ = the change in readiness for sponsor/claimant S/C from time $t-1$ to time t

$F_{S/C,t}$ = funding for sponsor/claimant S/C at time t

$F_{S/C,avg}$ = historical average funding for S/C.

The parameters to be estimated are the individual sponsor/claimant intercepts, $I_{S/C}$, and the slope, M. Given this equation, the break-even level of funding, denoted by $F^*_{S/C}$, can be found by setting $\Delta R_{S/C}$ = 0. The result is:

$$F^*_{S/C} = - \frac{I_{S/C}}{M} F_{S/C,avg} . \tag{2}$$

Notice that a sponsor/claimant's break-even funding level depends on both the slope and the intercept. Thus, the statistical significance of predicted break-even funding levels will depend on the significance of both slope and intercept. Although the estimated slope in Price Water-house's model is significant, table 1 shows that the sponsor/claimant intercepts are significant at a 5-percent level in only 13 of the 24 cases.

Another issue with the model's predictions is whether the predicted break-even funding level lies within the range of funding observed for that sponsor/claimant. Funding for sponsor 4/claimant 25 (logistics, NAVFAC) ranged between $131.6 and $136.3 million from 1984 through 1986.[1] The model, however, estimates that funding would have to be increased to $214.5 million to maintain constant readiness. This would represent an increase of 57 percent over the historical average funding level.

Predictions that lie outside the range of observed values in the sample must be treated with caution. Some sponsor/claimants have never, within the sample period, received funding levels that allowed them to keep readiness constant over time. There is obviously no direct infor-mation available in these cases on how much funding would have to be increased to maintain readiness. Table 2 shows that, for 13 of the 24 sponsor/claimants, the predicted break-even funding level lies out-side of the observed range of funding. It is argued in a following section that the relationship between funding and readiness cannot be

---

1. The funding variable used in the model is the sum of the current year's MRRP funding and the previous year's R/M MILCON funding. All funding amounts are expressed in millions of FY 1988 dollars.

linear over large changes in funding. If this is true, then the pre-
dicted break-even funding levels are even more suspect.

The intercept used to predict break-even funding is statistically
significant for 13 of 24 sponsor/claimants. Predicted break-even fund-
ing levels lie within the observed range of funding for 13 of
24 sponsor/claimants. Both tests are met in only 8 of the 24 cases.
The model's prediction that a 23-percent increase in funding would be
needed to maintain constant readiness must therefore be heavily quali-
fied. It would be preferable to make predictions closer to the observed
average funding levels and to present them with some indication of their
reliability.

Out-of-Sample Results

An examination of the out-of-sample predictions also demonstrates
the model's low level of accuracy at the sponsor/claimant level. Out-
of-sample predictions were made by deleting one of the years' data,
reestimating the model, and then determining predictions for the deleted
year. Each sponsor/claimant pair has three out-of-sample predictions.

Predicting the right direction of a change in readiness given a
certain level of funding would seem to be an important criterion in
deciding whether to use the model to allocate funds among sponsor/
claimants. The report claims that certain estimating techniques were
used to give added weight to the larger sponsor/claimants (in terms of
current plant value (CPV)) and thereby increase the model's accuracy
with respect to them. Therefore, the predicted direction of the change
in readiness was compared to the actual direction of change in readiness
for both the top five sponsor/claimants and the entire sample.

In predicting the right direction of a change in readiness, the
model predicts 24 incorrectly and 33 correctly, for an error rate of
42 percent (only nonzero predictions and observations were used in the
statistics). One would expect that a totally random forecasting proce-
dure (such as flipping a coin) would have an error rate of 50 percent.
The model performs a little better considering only the five highest CPV
sponsor/claimants. For this group, the model predicts the direction
correctly in five cases and incorrectly in eight cases for an error rate
of 38 percent.[1]

Rather than considering the accuracy of predicting the correct
direction of change, it is instructive to consider how well the model

---

1. There are 24 sponsor/claimants and three years for which tests can be
performed, for a total of 72 cases. Of these, in 15 cases either the
predicted or actual change in readiness is zero. Thus, there are
57 nonzero cases. For the five highest CPV sponsor/claimants, there are
a total of 15 cases, of which 13 are nonzero.

performs as compared to a much simpler model. The simplest model is one that states that readiness will be the same the next period as in the previous period irrespective of the level of funding. In other words, the change in readiness is always equal to zero. Two standard statistics were calculated in order to compare the two models: the root mean squared error (RMSE) and the mean absolute error (MAE). The root mean squared error is similar to the standard deviation but measures the dispersion of the forecast from the true value rather than the mean of the predictions. The mean absolute error is the average of the absolute value of the forecast errors. The MAE measures the average error on either side of the true value.

The RMSEs and MAEs for the full model and the simple model are given in table 3. The predicted changes were weighted by total sponsor/claimant CPV and the same statistics calculated in order to examine this alternative measure of readiness.

TABLE 3

COMPARISON BETWEEN A SIMPLE MODEL OF READINESS
AND THE FULL MODEL

Statistic	Full model	Simple model
Change		
RMSE	1.68	1.27
MAE	9.67	6.61
RMSE (top five)	0.38	0.27
MAE (top five)	5.11	3.91
CPV*Change		
RMSE	1.85	1.80
MAE	10.1	8.34
RMSE (top five)	1.49	1.59
MAE (top five)	21.8	21.3

As shown by table 3, the simple model always outperforms the full model, with the exception of the CPV-weighted RMSE for the top five sponsor/claimants. In some situations, especially for the MAE, the differences are substantial. Therefore, the use of the full model for predictions at the sponsor/claimant level can produce substantial errors.

At the sponsor/claimant level, the model produces results that may lead to a misallocation of resources, inaccurate forecasts of the direction of change in readiness, and performs only as well as, if not worse

than, a much simpler model. Therefore, the Price Waterhouse model should not be used to allocate funds at this level of aggregation.

SUGGESTED REVISIONS OF PRESENTATION

Although the presentation of the model in [1] is generally clear and complete, there are a few instances in which the interpretation of results may be misleading. First, the executive summary is too brief and presents the results of the model without sufficient qualification. Second, the results in the executive summary describe changes in readiness starting from break-even funding levels, and it would be preferable to give results starting from average funding levels. A third problem is that the model that is documented in [1] is not the model that was delivered to the Navy in a LOTUS spreadsheet. Finally, it is difficult to make inferences regarding the statistical significance of coefficients in the final model because the final model is the result of a lengthy specification search.

The Executive Summary

The executive summary reports two results. First, it states that "a 23 percent increase in combined MRRP and R/M MILCON funding (as compared to the average funding for 1984-1986) is required to overcome facility deterioration and maintain readiness at a constant level." This result comes from exhibit III-1 in [1] and was discussed in the preceding section on Predictions for Individual Sponsor/Claimants. The prediction is based on the individual intercept terms, over half of which are not significant. Also, the predicted break-even level of funding is outside of observed funding levels for over half of the sponsor/claimants. It would be difficult to calculate a confidence interval around this prediction because it is a function of 25 different coefficients. It is clear, however, that one should not place much confidence in the claim that 23-percent higher funding would prevent further deterioration of readiness. This result should not be presented as if it were fact, with no reference to its statistical validity, or to the difficulties with the underlying data.

The second result in the executive summary is that "Percentage increases from the break-even funding level result in proportional changes in readiness based on a factor of approximately 0.12." No reference is made to this result in the text, but Price Waterhouse analysts stated that it comes from the estimated slope in equation 1. If this is so, then the statement would be correct if it referred to increases from the average, rather than the break-even, level of funding (see appendix A on interpreting the model's coefficients).

Furthermore, since it is based on a single estimated coefficient, this result should be presented as a confidence interval rather than a point estimate. Using a confidence interval is preferable because it conveys some of the uncertainty that must be present with statistical

-9-

results. Rather than being approximately 0.12, the proportional change in readiness can be said to lie between 0.07 and 0.16 with 95-percent confidence.[1]

To illustrate how this coefficient could be interpreted, suppose that holding funding at the 1984 through 1986 average level is expected to cause a 2.5-percentage point decline in readiness each year. This is the prediction of the model in [1], as derived in this paper's appendix A. Thus, the 70.8-percent readiness achieved in 1986 would be expected to decline to 68.3 percent in 1987. If real funding were increased by 3 percent over the average level, then with 95-percent probability there would be less of a decline in readiness of from 3(.07) = 0.21 to 3(.16) = 0.48 percentage points. Predicted readiness in 1987 would then lie between 68.5 and 68.8 percent.

The last sentence in the executive summary, in which a 13-percent increase in funding is claimed to lead to a 1.2-percent decline in readiness, is incorrect. A 13-percent increase in funding is probably too large a change to have much confidence in its predicted effect on readiness. However, if it were a change from average funding levels, predicted readiness in 1987 would be between 69.2 and 70.4 percent with 95-percent probability. The point estimate would be 69.9 percent rather than the 69.6 percent suggested in the executive summary.[2]

In summary, the CNA study team believes the results in the executive summary should be qualified by references to their levels of statistical significance. Furthermore, there should be some discussion of the inherent difficulty of estimating a resources-to-readiness model using the BASEREP data. Finally, as is discussed in the section immediately following, it would be preferable to interpret the coefficients in terms of changes from average rather than break-even funding levels.

Average and Break-Even Funding Levels

A confusion between changes from average and break-even levels of funding is evident throughout the paper. There are two reasons why it would be preferable to interpret the coefficients in terms of changing funding from historical average levels. First, the problem of making

---

1. Equation 4 of appendix B in [1] gives the coefficient as 11.69 and the standard error as 2.44. The coefficient is divided by 100 to give the effect of a percentage change in funding.
2. The point estimate of 69.9 percent comes from multiplying the 13-percent increase in funding by 0.12 and adding the resulting 1.6-percentage point improvement to the 68.3-percent readiness expected in 1987 at average funding levels. Price Waterhouse's point estimate of 69.6 percent comes from decreasing 1986 readiness of 70.8 percent by 1.2 percentage points. It is assumed that they meant a 1.2-percentage point decline rather than a 1.2-percent decline in readiness.

forecasts outside of observed ranges of funding would be lessened.
Second, appendix A shows that changes in readiness starting from average
funding levels depend on the estimated slope, M. On the other hand,
changes in readiness starting from break-even funding levels depend on
the intercept terms, $I_{S/C}$. In the sponsor/claimant model, there are
24 of these intercepts and many of them have large variances. Thus,
changes in readiness from break-even funding levels are difficult to
calculate and are imprecise.

The preceding section on Predictions for Individual Sponsor/
Claimants points out that 13 of the 24 sponsor/claimants never experi-
enced the level of funding that the model predicts is necessary to
maintain readiness. Thus, even starting at the break-even funding level
implies a prediction outside the observed relationship between funding
and readiness. Increasing funding from the break-even level requires an
extrapolation even further beyond the limits of observed behavior.
Starting at the average level of funding in the data, however, means
that predictions at least start from a level of readiness that has been
experienced. One should be wary even in this case, however, of predict-
ing changes in readiness associated with large changes in funding. It
would be desirable to have the computerized version of the model print
warnings when predictions are beyond the limits of the estimation
sample.

Appendix A derives the following results regarding how the model's
coefficients can be interpreted. First, let ΔR^* be the change in
readiness that would result from holding funding constant at its histor-
ical average level (in this case, the average for 1984 through 1986).
For an individual sponsor/claimant, this change in readiness is given by

$$\Delta R^*_{S/C} = I_{S/C} + M .\qquad(3)$$

Aggregating over all sponsor/claimants, the model in [1] estimates
that $\Delta R^* = -2.5$. Then, if funding in period t is changed by x per-
cent from the average level, the resulting change in readiness is given
by

$$\Delta R_t = \Delta R^* + (0.01M)x .\qquad(4)$$

Since the estimated value of M is 11.7, if funding were increased
by 3 percent over the average level, a change in readiness of $-2.5 +$
$(0.117)3 = -2.1$ percentage points would be forecast. As has been
pointed out, it would be preferable to present this result using a
confidence interval rather than a point estimate.

Alternatively, let F^* be the level of funding that is required to
keep readiness constant (see equation 2). This funding level is 23 per-
cent higher than average funding for the model in [1]. Starting from
$F^*_{S/C}$, if funding changes by x percent in period t, an individual
sponsor/claimant will have a change in readiness of

$$\Delta R_{S/c,t} = \left(-0.011_{S/C}\right)x . \tag{5}$$

Aggregating over all the sponsor/claimants, the model in [1] predicts that an x-percent change in funding from the break-even level would cause a 0.14x-percentage point change in readiness. It is difficult to construct a confidence interval for this result because it is based on a weighted average of the 24 sponsor/claimant intercepts.

It is preferable to make forecasts using equation 4 rather than equation 5 for several reasons. First, in equation 4 it is not necessary to push funding up to the break-even level and then work backwards. Second, equation 4 uses the slope that is more likely to be statistically significant than all the intercept terms required in equation 5. Third, the computation is easier in equation 4 since only one estimated coefficient is required rather than 24. This also implies that it is easier to construct confidence intervals for the forecasts generated by equation 4.

Navywide Versus Sponsor/Claimant Models

Reference [1] reports the results of a model with different intercepts for each sponsor/claimant. This model could be used to allocate funds among sponsor/claimants, but since the accuracy of the model is so poor at this level, this action would not be recommended. When Price Waterhouse submitted a LOTUS model to the Navy, it was based not on the sponsor/claimant model, but on a Navywide model. That is, the model had been reestimated with just one intercept term. Although no documentation was made available of this model, it is possible to look at the LOTUS spreadsheet and find the estimated readiness equation:

$$\Delta R_{S/C,t} = -15.0 + 11.6 \, \frac{F_{S/C,t}}{F_{S/C,\ avg}} . \tag{6}$$

One recommendation is that the model documented in the report and the model delivered in the spreadsheet be the same. Furthermore, it is not clear that the Navywide model used in the spreadsheet is to be preferred to the sponsor/claimant model in the report.

Although the model is not accurate enough at the sponsor/claimant level to use the less aggregated results, it is still possible to use the sponsor/claimant model to make predictions at the Navywide level. Although the math would be more complicated, predictions could still be implemented on a spreadsheet without difficulty. Because the sponsor/claimant intercepts as a group add significantly to the explanatory power of the model, it is against standard practice to eliminate these

intercepts.[1] That is, a model with individual sponsor/claimant intercepts would be expected to make more accurate predictions at the Navywide level than would a model with only one intercept.

Furthermore, the Navywide model in the LOTUS spreadsheet generates more pessimistic predictions than does the model in the report. The preceding section on Navywide Readiness Predictions pointed out that the predictions of the Navywide model seem too pessimistic to be reasonable. Equations 3 and 6 above imply that the Navywide model estimates that holding funding constant at 1984 through 1986 averages will cause readiness to fall by 3.4 percentage points per year. This can be compared to the 2.5 yearly decline estimated by the sponsor/claimant model. The point estimate of 1987 readiness if funding were increased by 3 percent over average levels is 68.7 percent using the sponsor/claimant model. Using the Navywide model, this falls to 67.9 percent. This difference would be magnified as predictions are made out to 1994.

Reporting the Results of a Specification Search

The final model reported in [1] was the result of a lengthy specification search. That is, many preliminary regressions were run and the results of these regressions were used to choose a final model. This is a widely used technique, and some amount of specification search is virtually unavoidable. The statistics associated with a model arrived at by this method, however, must be used with caution. In particular, the standard errors of the coefficients will be biased and should not be used in tests of significance. To illustrate using an extreme example, suppose that one runs 100 different regressions trying to find a statistically significant relationship between funding and readiness. In the different regressions, different variables are included, different weighting schemes are used, and different functional forms are used. Finally, the one combination of all these factors is found that results in a coefficient on funding that is more than twice as large as its standard error. Reporting only this final regression and claiming that funding has a statistically significant effect on readiness would be misleading. Rather, this result and the range of possible coefficients should be reported.

Price Waterhouse should be commended for describing clearly the procedure used in their specification search, and for reporting many of their intermediate results. They do, however, base tests of statistical significance on the standard errors of the final model without even mentioning biases caused by specification search. Several of their intermediate results make it seem probable that in truth there is not a

---

1. No test statistic was reported for the joint significance of the intercept terms. Since 12 of the 24 are individually significant, however, it is assumed that they are significant as a group. However, this assumption should be tested.

statistically significant relationship between funding and readiness. At the very least, it seems likely that their final model overstates the effect of funding on readiness.

For example, if the data are weighted by CPV only, rather than by CPV and the number of units, the results are reportedly "similar but less significant." One suggested refinement to the estimation method made in a following section is to choose just one thing to weight by. Also, exhibit II-17 shows the results of two models. The first model measures readiness by the percentage of ratings that are C1 or C2, and the second model measures readiness by the percentage of CPV that is C1 or C2. The first model is adopted by Price Waterhouse because it shows a more significant relationship between funding and readiness. It is suggested in the section on Defining Readiness that follows, however, that the second measure of readiness may be preferred. As a final example, exhibit II-22 reports the results of adding variables such as age and usage to the model. Although none of the variables are significant, and thus are dropped, adding them does decrease the magnitude of the coefficient on funding and its level of significance. It is possible that omitting these other variables causes the importance of funding changes to be overstated because it is serving as a proxy for omitted variables.

The point here is that a final model may be chosen because the investigator is searching for a significant relationship between funding and readiness of a certain sign and magnitude. The results of this final model may say as much about the criteria used in the search as they do about the true relationship between funding and readiness. For this reason, the statistics associated with the final model should be reported with the proper qualifications and used with caution.

SUGGESTED REFINEMENTS OF THE MODEL

The preceding section discussed how the presentation of the existing Price Waterhouse model could be revised. It is the opinion of the CNA study team, however, that the existing model could be improved. There are problems both in the handling of the data and in the estimation methods used. This section will discuss these problems and suggest solutions. A detailed discussion of an alternative model that incorporates these suggestions is given in appendix B.

Data Problems

Some of the problems with the BASEREP data are insurmountable. The shortness of the sample period and the subjective nature of the readiness ratings cannot be changed. Price Waterhouse has already done a good job of handling other problems with the data. There are several problems, however, that are not addressed sufficiently in [1]. First, the issue of how readiness is to be measured must be decided before any meaningful model can be developed. Second, some additional independent

variables are suggested for inclusion in the model. Finally, the implications of the small sample size for the type of model that should be estimated are discussed.

Defining Readiness

The purpose of the Price Waterhouse modeling effort is to relate funding to facility condition readiness. One would like an estimate of how changes in funding might affect the ability of facilities to fulfill the requirements placed on them. It is obvious that for the model to be successful, it must be based on a meaningful measure of readiness. Whether a measure of readiness is meaningful or not must be decided by the users of the model. This issue cannot be decided by any statistical test.

The readiness measure used in the Price Waterhouse model is the percentage of ratings that are either C1 or C2. For each activity, ratings are given in all appropriate mission categories. Using the percentage of C1 and C2 ratings as a measure of readiness gives equal weight to port operations in Long Beach and fire protection in a Reserve Center. With this measure of readiness, it is hard to know the gravity of the decline in readiness depicted in figure 1. It is possible that the 25 percent of C3 and C4 ratings in 1983 included the most important missions of the largest installations. On the other hand, the 54 percent of C3 and C4 ratings projected for 1994 could include only the less important missions in the smaller installations.

Alternative readiness measures would attempt to weight the ratings by measures of the size and importance of the mission and activity. For example, the current plant value (CPV) associated with all missions that received C1 or C2 ratings could be totalled. The CPV_1 that is ready could then be expressed as a percentage of total CPV.[1] This is referred to in the Price Waterhouse report as a CPV-weighted average readiness rating.

Another possibility for weighting the ratings is to use Shore Facilities Life Extension Program (SFLEP) priority categories. The SFLEP assigns high, medium, and low priorities by investment categories. This weighting could be combined with the CPV weighting to produce a measure that would indicate, for example, what percentage of CPV is ready in high-priority categories.

Any weighting scheme introduces complications. The reliability of the CPV numbers in the Navy Facility Assets Data Base (NFADB) has been questioned by OP-44 and Price Waterhouse. The SFLEP priorities are assigned to investment categories, which do not correspond perfectly

---

1. An alternative would be to have commanding officers report readiness as the percentage of CPV in a certain mission that is C1, C2, C3, or C4.

into the mission categories used in BASEREP. If a model that simply
predicts the percentage of C1 and C2 ratings provides the information
that the Navy needs to allocate MRRP funds, then the complications of
weighted readiness measures can be avoided. If, however, a more sophis-
ticated measure of readiness is required, then these complications must
be addressed.

Price Waterhouse dismisses CPV-weighted average ratings because
they are more variable and result in a poorer fitting model. It is
wrong, however, to decide what measure of readiness to use based on
statistical tests. This question can be decided only on the basis of
what the Navy needs to know to correctly allocate funds. It is
necessarily undesirable to have more variation in a readiness measure.

Suppose that the goal is to allocate funds dependent on the CPV of
sponsor/claimants. Then a CPV-weighted readiness measure is the correct
measure, and moving to the unweighted measure means discarding variation
in readiness that the model does not explain well. The resulting
increase in the goodness of fit and appearance of logic in the model
would be entirely spurious.

Before the modeling effort can proceed, the proper measure of
readiness must be decided upon. This decision can be made only by the
people who intend to use the model to make resource-allocation deci-
sions.

Measurement Error

The readiness data collected from 1983 through 1986 are highly
subjective. Facilities are rated by an individual from the particular
facility and approved by the commanding officer. Both the individual
preparing the report and the commanding officer may have incentives to
be biased in reporting readiness. For example, low ratings may be
perceived as a method to justify a request for additional funding. Low
ratings at the beginning of a tour followed by progressively higher
ratings throughout the tour may be an indication of improvement due to
the management of the facility. Further, changes in personnel may have
a significant effect on reported readiness, while actual readiness
remains unchanged.

These effects may cause actual readiness to have a low covariance
with reported readiness. A model relating reported readiness to funding
may be different from a model relating actual readiness to funding. A
possible correction for this problem in the model may be the inclusion
of a variable representing a change in the personnel preparing or
approving the report.

Other effects may also cause actual readiness to be different from
reported readiness. In the earlier years' data, the users of the facil-
ities often did not have input into the rating process. Rather, the

owners of the facilities were responsible for reporting the readiness of a facility. The owners could differ from the users of the facility if the facilities were leased from another organization. In some situations, the owners did not consult with their tenants in preparing the readiness reports. This oversight has been corrected in the latest readiness reports, which now require consultation with tenants. The addition of a variable representing the percent of leased facilities should be included in the model to determine if the reporting differs when more facilities are leased.

An increase in the number of missions covered by the reporting system and the transfer of facilities between sponsors may cause the relationship between readiness and funding to change over time. The total number of reports increased 23 percent from 1984 through 1985 from 1,243 to 1,528. From 1985 through 1986, however, the number of reports stayed relatively constant, from 1,528 to 1,542. Because it appears that there may be a different population for 1984 as compared to 1985 and 1986, estimating single parameters that do not change over the years may not be valid. Similarly, the transfer of facilities from one sponsor/claimant to another may also cause the estimated parameter to differ over the years. The magnitude of these effects may be determined by constructing and estimating the model with a consistent data base that has the same reports for each sponsor/claimant over all four years. If this is not possible, the data should be tested, as described in appendix C, to determine if there are different populations.

Small Sample Size

Current data are provided for four years. When the lagged values are constructed, only three years of data are available for the estimation process. The Price Waterhouse model estimates a separate coefficient for each of the sponsor/claimants. Therefore, three observations are used to estimate these sponsor/claimant-varying coefficients. The low accuracy obtained by this process is shown by the relatively high standard errors reported for these variables. Possible solutions to this problem are to estimate fewer coefficients by grouping some sponsor/claimants together or to wait until more data are available.

Estimation Problems

Treatment of Time-Series/Cross-Section Data

Although the Price Waterhouse model in general has the virtue of simplicity, its treatment of time-series/cross-section effects is unnecessarily complicated. To illustrate, begin with a simple time-series/cross-section model:

$$y_{it} = a + bx_{it} + u_{it} . \tag{7}$$

-17-

In this model, there are observations for individuals, i, over a series of time periods, t. Some variable y is assumed to be a linear function of the independent variable x and a random error u. The parameters of the model are a and b. In the facility condition application, the individuals are the sponsor/claimants, y is readiness, and x is funding. Suppose that each sponsor/claimant can be assumed to have some fixed effect on readiness. That is, due to some omitted variable such as historical funding levels, or the age of the facilities, sponsor/claimant i will have a higher level of readiness than sponsor/claimant j. This difference in readiness is independent of funding levels and constant over time. In this case, it is appropriate to estimate a model with a different intercept for each sponsor/claimant:

$$y_{it} = a_i + bx_{it} + u_{it} \; .$$ (8)

The model in equation 8 is inappropriate if the fixed sponsor/claimant effects are believed to be correlated with the independent variable (for example, if sponsor/claimants who have historically been underfunded also tend to be underfunded in the current period; also, if some sponsor/claimants have better facility managers who both keep readiness higher and succeed in winning higher levels of funding). If such a correlation is expected, then a first-difference model is commonly estimated:

$$\left(y_{it} - y_{i,t-1}\right) = b\left(x_{it} - x_{i,t-1}\right) + u_{it} \; .$$ (9)

A variation on the first-difference model would be to use ratios rather than differences, and to take the ratios relative to the average over time for the sponsor/claimant:

$$\left(y_{it}/y_{i,avg}\right) = b\left(x_{it}/x_{i,avg}\right) + u_{it} \; .$$ (10)

The model estimated by Price Waterhouse is a combination of the models given in equations 8, 9, and 10

$$\left(y_{it} - y_{i,t-1}\right) = a_i + b\left(x_{it}/x_{i,avg}\right) + u_{it} \; .$$ (11)

The model becomes even more complicated when other independent variables are added to the model, apparently in first-difference form.

It is recommended that one or another of the standard time-series/ cross-section models be adopted. That is, either sponsor/claimant intercepts be included, or variables be expressed as differences (or ratios) from past values (or to average values). Combining the different models may produce unexpected statistical results. Furthermore, the coefficients in their hybrid model are difficult to interpret. The slope term does not give the change in readiness implied by a certain change in funding. Rather, it gives the change in the change in readiness when funding is changed from its historical average level.

Specifying a Stock/Flow Problem

In section II.B.3 on pages II-28 to II-32 of [1], there is a discussion of what Price Waterhouse refers to as level and change models. The essence of this discussion is whether the true value of A equals one in the following equation:

$$R_t = A \cdot R_{t-1} + I_{S/C} + M\left(\frac{F_{S/C,t}}{F_{S/C,avg}}\right) + e_t \ . \tag{12}$$

If the true value of A is one, then it is proper to estimate the equation with the change in readiness, $\Delta R = R_t - R_{t-1}$, on the left-hand side. This is referred to in [1] as a change model. If it cannot be assumed that the true value of A equals one, then R_{t-1} should remain on the right-hand side, and A will be a parameter to be estimated. This is referred to as a level model.

Rather than discussing whether the true value of A equals one, however, the report begins the discussion by assuming that the true value is one. There are both theoretical and empirical reasons to believe, however, that A does not equal one. First, readiness is a stock, and funding for maintenance and repair is a flow. Readiness is defined to be the condition of a facility at a certain point in time. Funding, however, is an amount per unit time, for example, per year. Stocks usually depreciate at a certain percentage per year rather than at a fixed amount as specified by the model when it combines the difference and fixed-effect model. At low levels of readiness, a fixed-amount decrease in readiness given constant funding may cause readiness to become negative. A more reasonable assumption would be that readiness in one year will be some fraction of the previous year's readiness, or that $A < 1$.

Furthermore, Price Waterhouse provided the study team with some additional regression results in which R_{t-1} was placed on the right-hand side. In many of these regressions, the test of the null hypothesis that A equals one is rejected. In all cases, the estimated value of A is no greater than one. Thus, there is empirical support for the theoretical presumption that A is less than one.

-19-

Autocorrelation and Lagged Readiness

After assuming that A equals one, the report goes on to argue
that making lagged readiness an independent variable will create prob-
lems with autocorrelation and errors in variables. Both of these asser-
tions are incorrect. A model that uses time-series data very likely
will have autocorrelated errors. This is particularly true when there
are many omitted variables, as there are in the Price Waterhouse model.
Autocorrelation is a problem that should have been tested for and cor-
rected. Having lagged readiness as an independent variable would make
these tests and corrections slightly more complicated, but it would not
in itself cause autocorrelated errors.

A similar argument is put forth regarding errors in variables. The
BASEREP readiness measures are not perfect measures of true facility
condition readiness. In other words, the readiness measure suffers from
measurement error. In regression models, measurement error in indepen-
dent variables poses greater problems than does measurement error in the
dependent variable.[1] Therefore, Price Waterhouse argues that it is
better to have lagged readiness as a dependent variable.

If the true value of A is not one, however, moving R_{t-1} to the
left-hand side will not solve measurement-error problems. At best, it
can be thought of as trading errors-in-variables bias for specification-
error bias. Moreover, there are methods available to correct for
errors-in-variables problems.

In summary, if the true value of A is not one, then the problems
of autocorrelation and errors in variables cannot be avoided by
estimating a "change" model. Such a model would be a misspecification
of the correct model and will lead to biased parameter estimates. If
A does not equal one, then lagged readiness must be on the right-hand
side. The problems of autocorrelation and errors in variables cannot be
assumed away, no matter how inconvenient they may be.

Functional Form

The Price Waterhouse model uses a linear functional form. It is
argued that even if the true readiness function is not linear, a linear
relationship can approximate a curve within a limited region. This is
true, but the results of the estimation are not used to predict changes
in readiness within a limited region. For this reason, and because

---

1. It is not true, however, as Price Waterhouse implies, that
measurement error in the dependent variable creates no problems.
Estimates of coefficients will be unbiased, but estimates of standard
errors using these data will be biased away from the errors of the
actual model. Estimates that would be obtained from using correctly
measured data would be unbiased.

there are theoretical reasons to believe that the relationship between funding and readiness is not linear, the CNA study team believes that a nonlinear functional form should be used. This is an important point because the linear functional form may be responsible for the model's overly pessimistic predictions.

The dependent variable in the Price Waterhouse model is the change in the percentage of ratings that are C1 or C2. This variable must range between -100 and +100. For this reason, the curve relating funding to readiness should approach these values asymptotically. Assume that the true relationship between the change in readiness and funding is the curve labeled "assumed true relationship" in figure 2. (Appendix B discusses this functional form in more detail.) The curve has the properties discussed above. The sample that Price Waterhouse uses is concentrated below the axis where the change in readiness is zero. Appendix A in [1] shows that, out of 69 observations in the sample, 36 had a decrease in readiness, 12 had no change, and 21 had an increase. The scatter of points in figure 2 is a loose representation of the actual observations on funding and readiness changes in the sample. A linear approximation of this scatter of points would resemble the line labeled "estimated relationship" in figure 2.
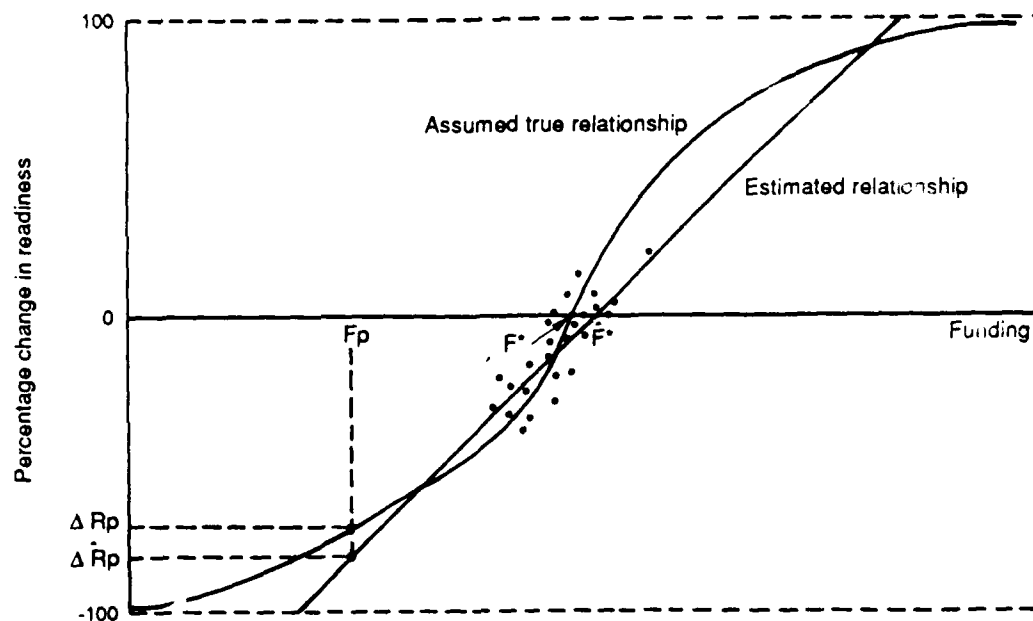

FIG. 2: FUNCTIONAL FORM

-21-

There are several things to notice about the true and estimated relationships depicted in figure 2. First, for many values of funding, the estimated change in readiness lies below the true change. That is, the estimated relationship would be biased toward making pessimistic predictions. Second, consider the level of funding necessary to keep readiness constant. The linear approximation predicts a level, F^*, that is in excess of the true break-even level, F^*. Finally, F_p represents a large cut in funding, such as in the out-years of the President's budget. The predicted decrease in readiness using the linear approximation would be greater than the true decrease in readiness.

If the true relationship between readiness and funding is nonlinear and a linear relationship is estimated, predictions based on the linear model will be biased for large changes in funding. In particular, if the sample heavily represents negative changes in readiness, the break-even funding level may be overestimated. Predicting changes in readiness related to funding levels that lie within the sample range may involve some bias. Predictions outside of the sample range are always unreliable, but will be even more so if an incorrect functional form that does not have known properties is used.

Use of Weighted Least Squares

In the report, the explanation given for using weighted least squares is not entirely correct. Although weighted least squares does give added emphasis to certain observations as stated in [1], it is important to realize that this is not detrimental to the other observations and is not the usual reason for using this method of estimation. Specifically, the estimates obtained by this method are not biased. Estimates of all values of readiness should be closer to their true value.

Weighted least squares is the method usually used to correct for the unequal variance that may occur across observations. The unequal variance violates a basic assumption of ordinary least squares. The correction produces estimates of the parameters of a model that have a lower variance than ordinary least squares (that is, they are more exact). Moreover, the procedure corrects for the biased estimates of the variance of the parameters produced from least-squares estimation. The biased estimates of the variances cause incorrect test statistics to be calculated.

It is not surprising that the variance decreases with the number of Unit Identification Codes (UICs) or CPV. Both are proxies for the total number of ratings. Since each particular rating can take on only one of two values (C1 and C2 ratings are given one value and C3 and C4 ratings are given another value), the proportion of positive ratings has a distribution that is probably very similar to a cumulative binomial distribution. Therefore, the variance of the proportion of positive

-22-

ratings is inversely proportional to the number of ratings in each observation. However, the variance is also related to the proportion of positive ratings. A correction for the change in variance over the observations caused by this relation should also be performed.

CONCLUSIONS AND RECOMMENDATIONS

The model developed by Price Waterhouse does not produce reasonable results at the Navywide level and is very inaccurate at the sponsor/ claimant level. For these reasons, the CNA study team recommends that the model not be used in its present form. It is possible that no useful model can be developed from the existing data due to the data's subjectivity and variability. However, the study team recommends that other approaches should be tried before the decision is made on whether to use a formal model for the allocation of resources.

Based on a review of [1], the CNA study team recommends that the following tasks be performed:

- The Navy should decide on a definition of readiness that incorporates a measure of the importance of the different missions.

- The data for 1987 should be tested for compatibility with the existing data as described in appendix C and combined with the existing data if compatibility is confirmed by the tests.

- The model as specified in appendix B should be estimated, tested to determine if it is consistent with the theory on which it is based, have future values forecast, and have the reasonableness of those forecasts determined.

- Because the sample size is limited, combining some of the parameters, especially the dummy variables, should be explored.

- Throughout the text, but especially in the executive summary, the variability of the results should be documented and discussed.

- The report should document the model actually delivered to the Navy in a LOTUS spreadsheet.

- Once a new report is delivered to the Navy, a decision should be made on whether a formal model is appropriate given the existing data. If a formal model is not used, other approaches for allocating resources should be developed.

REFERENCES

[1] Price Waterhouse, Contract No. N00600-86-D-3869, Delivery Order 2, *Development of an Analytical Model Relating MRRP Resources to Facility Condition Readiness*, Aug 1987

[2] CNA Research Memorandum 86-169, *Issues in the Measurement and Modeling of Shore Base Facility Readiness*, by Henry L. Eskew, Aug 1986 (27860169)[1]

[3] CNA Research Memorandum 86-267, *A Study of Selected Issues in Military Construction and Base Operating Support*, by Henry L. Eskew, Bobby Jackson, and Joseph S. Domin, Dec 1986 (27860267)

[4] Price Waterhouse, Contract No. N00600-86-D-3869, Delivery Order 1, *Revision of BASEREP To Include Objective Readiness Criteria for Facilities Condition*, Dec 1986

---

1. The numbers in parentheses are internal CNA control numbers.

APPENDIX A

INTERPRETING COEFFICIENTS IN THE PRICE WATERHOUSE MODEL

APPENDIX A

INTERPRETING COEFFICIENTS IN THE PRICE WATERHOUSE MODEL

Interpreting the coefficients in the Price Waterhouse model is not straightforward because of the way the variables are specified. The model is given by

$$\Delta R_{S/C,t} = I_{S/C} + M\left(F_{S/C,t}/F_{S/C,avg}\right) . \tag{A-1}$$

All of the variables are defined following equation 1 in the text of this paper.[1] One would like to use the estimated values of the coefficients to make statements about how readiness will change given a change in funding. The slope term, M, however, gives the change in the change in readiness given a change in funding relative to average funding. This can be expressed as:

$$M = \frac{\Delta\left(\Delta R_{S/C,t}\right)}{\Delta\left(F_{S/C,t}/F_{S/C,avg}\right)} = \frac{\Delta R_{S/C,t} - \Delta R_{S/C,0}}{\dfrac{F_{S/C,t}}{F_{S/C,avg}} - \dfrac{F_{S/C,0}}{F_{S/C,avg}}} , \tag{A-2}$$

where $\Delta R_{S/C,0}$ is the change in readiness in some initial period, and $F_{S/C,0}$ is the level of funding in this initial period.

The coefficients can be interpreted more easily by setting funding in the initial period at certain levels. One set of interpretations arises if $F_{S/C,0}$ is set equal to the average funding level; another set arises if $F_{S/C,0}$ is set equal to the break-even funding level.

INITIAL FUNDING AT AVERAGE FUNDING LEVELS

Let $\Delta R^*_{S/C}$ be the predicted change in readiness when funding is held at average levels. If $F_{S/C,0} = F_{S/C,avg}$, it follows from equation A-1 that:

$$\Delta R^*_{S/C} = I_{S/C} + M . \tag{A-3}$$

---

1. The results in this section hold for a Navywide model as well. The only differences for the Navywide model are that the sponsor/claimant intercepts are replaced by a single intercept, I, and there is no subscript on the change in readiness.

Using this formula and the coefficients from equation 4 in [1]
appendix B, table A-1 predicts readiness in 1987. For these predictions
it is assumed that base readiness is at 1986 levels and that 1987 fund-
ing equals the average of 1984 through 1986 funding. Predicted readi-
ness in 1987 would fall to 67.7 percent, or by 2.5 percentage points
from the 1986 level of 70.2 percent. For the Navywide estimated model
given by equation 6 in the text, equation A-3 would predict a decrease
in readiness of 3.4 percent with average funding.

Now suppose that funding begins at the average level and is then
changed by x percent in the next period. Equation A-2 would reduce
to:

$$M = \frac{\Delta R_{S/C,t} - \left(I_{S/C} + M\right)}{1.01x - 1} \; . \tag{A-4}$$

Solving for the change in readiness results in:

$$\Delta R_{S/C,t} = I_{S/C} + (1.01x)M = \Delta R^*_{S/C} + (0.01M)x \; . \tag{A-5}$$

Holding funding at average levels would result in an estimated
2.5-percentage-point decline in readiness each year. The slope, M,
tells how this decline would be modified if funding were changed from
the average level. In particular, an x-percent change in funding would
cause a $(0.01M)x$-percentage-point change in the estimated decline in
readiness. For example, when $M = 11.7$, a 3-percent increase in funding
from average levels would cause the decline in readiness to be 0.4 per-
centage points less than -2.5 percent. Thus, the model would predict a
2.1-percent decline.

INITIAL FUNDING AT BREAK-EVEN FUNDING LEVELS

Let $F^*_{S/C}$ be the level of funding such that $\Delta R_{S/C,t} = 0$. From
equation A-1, it follows that:

$$F^*_{S/C} = -\left(I_{S/C}/M\right)F_{S/C,avg} \; . \tag{A-6}$$

Suppose that initial funding and readiness in equation A-2 are set at
these levels, and then funding is increased by x percent. Equation A-2
becomes:

$$M = \frac{\Delta R_{S/C,t} - 0}{-\left(1 + 0.01x\right)\left(I_{S/C}/M\right) + \left(I_{S/C}/M\right)} \; . \tag{A-7}$$

A-2

TABLE A-1

CHANGE IN READINESS AT AVERAGE FUNDING LEVELS

Claimant	Sponsor	Actual readiness in 1986[a]	$\Delta R*$[b]	Predicted readiness in 1987[c]	Number of ratings[d]	Predicted number of ready ratings[e]
11	1	82.8	-0.7	82.1	29	23.8
11	10	64.3	2.3	66.6	14	9.3
18	27	86.0	-2.9	83.1	150	124.7
23	4	84.3	-0.7	83.6	51	42.6
25	4	60.9	-6.6	54.3	46	25.0
30	2	100.0	0.9	100.0	21	21.0
60	1	72.7	-20.5	52.2	22	11.5
60	2	59.1	-3.1	56.0	22	12.3
60	3	75.0	-0.2	74.8	108	80.8
60	5	57.3	-5.6	51.7	171	88.4
60	16	100.0	9.0	100.0	16	16.0
61	2	85.7	1.2	86.9	21	18.2
61	3	60.0	-1.0	59.0	15	8.9
61	4	47.1	-8.5	38.6	17	6.6
61	5	89.6	4.3	93.8	48	45.0
62	1	52.5	-0.7	51.8	59	30.6
62	5	65.0	0.4	65.4	120	78.5
62	16	40.0	-16.8	23.2	10	2.3
70	2	73.9	-6.1	67.8	46	31.2
70	3	61.8	-2.7	59.1	110	65.0
70	4	73.8	1.4	75.3	88	66.3
70	5	66.2	-3.4	62.8	213	133.8
72	4	46.7	-13.5	33.2	15	5.0
72	5	75.4	-0.4	75.0	130	97.5
Navywide		70.2	-2.5	67.7	1,542	1,044.3

a. Actual readiness in 1986 and the number of ratings in 1986 are taken from appendix A of the Price Waterhouse report.

b. The predicted change in readiness at average funding levels, $\Delta R*$, equals $I_{S/C} + M$. These values are taken from equation 4 in appendix B of the Price Waterhouse report.

c. Predicted readiness in 1987 is the sum of actual 1986 readiness and $\Delta R*$.

d. The predicted number of ready ratings is the percentage of predicted 1987 readiness times the number of ratings.

e. The Navywide forecasts are determined as follows: predicted percentage readiness in 1987 is the total predicted ready ratings, 1,044.2, as a percent of total ratings, 1,542; $\Delta R*$ is predicted 1987 readiness, 67.7, from the 1986 readiness, 70.2.

Solving for the change in readiness results in:

$$\Delta R_{S/C,t} = (-0.01 I_{S/C})x \qquad .$$ (A-8)

Starting from the break-even funding level, then, the change in readiness does not depend on the slope but rather on the intercept terms. For the sponsor/claimant model, the Navywide change in readiness would be a function of all 24 sponsor/claimant intercepts. A weighted average of the estimated intercepts from the Price Waterhouse model is -14.0. The weights used for this calculation were the number of ratings for each sponsor/claimant. Using this result, a 10-percent fall in funding from the break-even level would cause a 1.4-percentage-point decline in readiness. In the Navywide model, where the estimated value of the intercept is -15.0, the same decrease in funding would be predicted to cause a 1.5-percentage-point decline in readiness.

APPENDIX B

AN ALTERNATIVE READINESS MODEL

APPENDIX B

AN ALTERNATIVE READINESS MODEL

In developing an analytic model that relates MRRP and MILCON resources to shore base facility condition readiness, it is important to develop a model that is based on a priori knowledge and economic theory relevant to the relationship. There are many reasons why this is important (see [B-1], chapter 2.) For example, if readiness is defined to be the percent of missions that are ready, then the relationship cannot be less than zero or greater than one. For small changes in the variables, a simple linear relationship would not present problems in estimating the relationship given the zero/one restriction. However, with large variations, as are seen in the readiness data, assuming a linear relationship with the zero/one restriction can cause substantial errors in the estimated relationship and forecasts for values outside of the sample.

This section develops a model of shore base facility condition readiness as a function of funding and other variables considering the economic theory and a priori knowledge about the relationship. The development is based on [B-2] and [B-3].

The interest in this study is in whether a particular physical facility is "ready", that is, able to carry out its mission. Readiness is, therefore, dependent on the physical condition of that facility. The physical condition can be measured by the total plant value[1] of that facility. With respect to total plant value, the economics' literature in investment theory has developed models to explain how facilities deteriorate and are improved over time through operation and maintenance expenditures or through further investment. Such models can be applied here.

Following investment theory, the total plant value (TPV) for a sponsor/claimant S/C in period t is a function of the total plant value at the end of the last period, the depreciation[2] that has occurred

---

1. The total plant value (TPV) of the facility is defined as the value of the facility to the Navy, similar to the value of physical capital in the private sector. TPV is not the same as CPV, as currently measured, as CPV does not take into account depreciation of the facility. In the private sector, the market places a value on physical capital; however, here the Navy must decide on the value. Since this variable is eliminated before the estimation process, it does not need to be measured.

2. Depreciation is defined as the actual physical deterioration of the facility. The measure of readiness used in this analysis is a measure of the condition of the facility and not the ability of the facility to

in the total plant value in the last period, and the new funding (*Fund*) for investment in improving the facility during the period. The parameters may differ across sponsor/claimants. Therefore, the parameters may be subscripted by S/C to refer to the particular sponsor/claimant pair to which they apply.[1] The relationship for the total plant value is given in the following equation:

$$TPV_{S/C, t} = TPV_{S/C, t-1} - a_{S/C}TPV_{S/C, t-1} + bFund_{S/C, t} + e_{S/C, t} . \qquad (B-1)$$

Readiness is dependent not on the absolute total plant value, however, but rather on the relative value to the original construction cost[2] (OCC) of the facility. The TPV at any given period of time may be greater than or less than OCC, dependent on the level of maintenance and depreciation, or even negative if the facility needs to be demolished. Therefore, the ratio is not limited to range between zero and one.

Further, there are other variables that may influence the reported value of readiness. An important factor may be the commanding officer of the facility or the survey respondent. The reporting of readiness, especially in the past, has been very subjective. A change in the individual making the subjective judgments may have a significant effect on the reported value of readiness.

Since the definition of readiness used here constrains the values it may obtain to be between zero and one, a linear form is not appropriate for large changes. A logistic functional form is assumed. The form is an *S*-shaped curve with an upper and lower asymptote as depicted in figure B-1. As the function approaches the lower asymptote, a greater reduction in the independent variable is needed for a given reduction in the dependent variable. Conversely, as the function approaches the

---

meet its mission. Similarly, the measure of funding is for maintenance of the facility rather than expansion of the facility or for updating the facility for new missions. Therefore, other traditional definitions of depreciation that may also involve obsolescence are not applicable. The rate of depreciation may be dependent on the usage of the facility and its age. The estimation of a model incorporating this dependence is difficult (see [B-4]).

1. In the following development of the model, it will be assumed that the depreciation term *a* will vary across sponsor/claimant pairs since it may vary with the usage of the facility and its age. Due to a limited sample size, not all parameters can vary. Whether a particular parameter varies across sponsor/claimants should be statistically tested.

2. The current plant value (CPV), as measured by the Navy, is almost identical to the definition of OCC used in this paper. Therefore, CPV can be used in the estimation process for OCC.

upper asymptote, a greater increase in the independent variable is
needed for a given increase in the dependent variable. The general form
of the logistic function is given in equation B-2. If u is greater
than s in equation B-2, then the upper asymptote is u and the lower
asymptote is s:

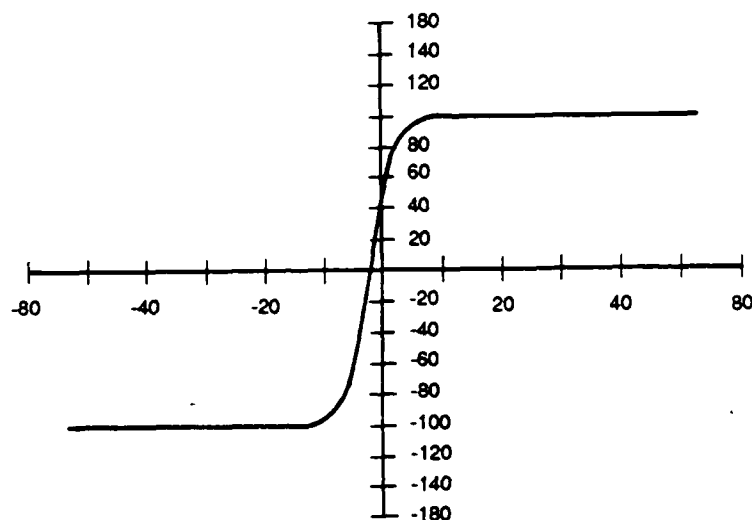$$ln \frac{Y - s}{u - Y} = V + wX \ . \tag{B-2}$$

FIG. B-1: LOGISTIC FUNCTIONAL FORM

The logistic form is ideally suited for a functional form relating readiness to other variables.[1] Since the readiness measure was constructed to range only between zero and one, the lower asymptote, s, should be constrained to have a value of zero, and the upper asymptote, u, should be constrained to have a value of one. Parameters represented by w (w could be a vector of parameters and X a matrix of variables) determine the response of the logistic form of readiness to changes in the independent variables. Due to the form of the function, as readiness approaches one, larger increases in the independent variables are needed to obtain a given incremental increase in readiness. Conversely, as readiness becomes close to zero, larger reductions in the independent variables are needed to decrease readiness by a given amount.

As in the total plant value equation, the parameters in the readiness relationship may differ. Therefore, the parameters may be subscripted by S/C to refer to the particular sponsor/claimant pair to which they apply. Equation B-3 specifies the assumed relationship between readiness and the independent variables consisting of the ratio between TPV and CCC, and a dummy variable representing personnel changes (DPC):

$$\ln \frac{R_{S/C,t}}{1 - R_{S/C,t}} = c + d \frac{TPV_{S/C,t}}{OCC_{S/C}} + fDPC_{S/C,t} + r_{S/C,t} . \qquad (B-3)$$

Equations B-1 and B-3 form a recursive system in that equation B-1 does not contain more than one dependent variable that is in equation B-3, and there is no correlation between the error terms in the two equations. Substituting equation B-1 into equation B-3 and eliminating the total plant value variable by subtracting lagged logistic readiness, yields equation B-4:

$$\ln \frac{R_{S/C,t}}{1 - R_{S/C,t}} = ca_{S/C} + \left(1 - a_{S/C}\right) \ln \frac{R_{S/C,t-1}}{1 - R_{S/C,t-1}} + \frac{db}{OCC_{S/C}} Fund_{S/C,t}$$
$$+ f\left(DPC_{S/C,t} - \left[1 - a_{S/C}\right]DPC_{S/C,t-1}\right) + n_{S/C,t} . \qquad (B-4)$$

The error term, $n_{S/C,t}$, in equation B-4 can be expressed in terms of the error terms in equations B-1 and B-3 as:

$$n_{S/C,t} = \frac{d}{OCC_{S/C}} e_{S/C,t} + r_{S/C,t-1} . \qquad (B-5)$$

_____

1. Other functional forms are possible, but the logit formulation is the most commonly used function relating percentages to other variables. For a generalization of this form, see [B-5] and [B-6].

Since the new error term is a linear combination of the other terms, if the previous terms were normally distributed with mean zero, then the error term in equation B-4 is normally distributed with mean zero. However, since the error term in equation B-4 contains a lagged error term from equation B-3, there may be autocorrelation of error terms in equation B-4.

Procedures for Estimating the Model

Estimation of the model is not straightforward because of the constraints on the parameters in the model. The estimation procedure involves construction of the variables; estimation; testing of the assumptions, constraints and error structure; and, reestimation correcting for the identified problems. The procedure is described as follows.

Construction of the Variables

Most of the variables in the model have been constructed for use in previous models. The only exceptions are the logistic form for readiness and lagged readiness and the dummy variable for personnel changes.

The particular form of the personnel dummy must be explored. The nominal definition would be that the variable takes on the value one for a particular commanding officer and the value zero otherwise. However, some sponsor/claimant pairs may consist of more than one facility and would have more than one commanding officer. The precise form of the dummy variable should be determined empirically. One alternative may be that the dummy variable would take on the value one for a group of commanding officers and the value zero otherwise. Another alternative may be that the dummy variable could represent a metric such as the percent of missions whose commanding officers remain the same.

Estimation

Initial estimation of the model will assume that there is no autocorrelation between error terms. This assumption simplifies the estimation procedure considerably and allows a test to be conducted to determine if the correlation is strong enough to warrant correction of the problem.

The equation that will be estimated, equation B-4, is repeated here for reference:

$$\ln\frac{R_{S/C,t}}{1 - R_{S/C,t}} = ca_{S/C} + \left(1 - a_{S/C}\right)\ln\frac{R_{S/C,t-1}}{1 - R_{S/C,t-1}} + \frac{db}{OCC_{S/C}} Fund_{S/C,t}$$
$$+ f\left(DPC_{S/C,t} - \left[1 - a_{S/C}\right]DPC_{S/C,t-1}\right) + n_{S/C,t} . \qquad (B-4)$$

Although it may appear that there are a large number of parameters, closer inspection reveals that restrictions on the parameters limit their number. A suggested method to estimate the model is to use the SYSNLIN procedure in SAS (see [B-7]). However, rather than estimating equation B-4, equation B-6 should be estimated using the RESTRICT statement to specify that $a_{S/C} = g_{S/C} = h_{S/C}$.

$$ln\frac{R_{S/C,t}}{1 - R_{S/C,t}} = ca_{S/C} + \left(1 - g_{S/C}\right)ln\frac{R_{S/C,t-1}}{1 - R_{S/C,t-1}} + \frac{db}{OCC_{S/C}}Fund_{S/C,t}$$
$$+ f\left(DPC_{S/C,t} - \left[1 - h_{S/C}\right]DPC_{S/C,t-1}\right) + n_{S/C,t} \quad (B-6)$$

The advantage of this method is that the test statistic for testing whether the restrictions are valid are printed, and the tests, as described in the SAS manual [B-7], are straightforward.

Testing of Constraints and Error Structure

After the initial estimation of the model, the assumptions and constraints of the model should be tested to determine if they are valid. The tests of $a_{S/C} = g_{S/C} = h_{S/C}$, as mentioned previously, would determine if the restrictions are valid. Tests of $a_{S/C} = a_{S/C}$ (for $i \neq j$) would determine if depreciation of the facilities is the same over all sponsor/claimants. Using one depreciation rate would considerably simplify the model. Finally, tests should be performed on other parameters to determine if they should vary across sponsor/claimants rather than $a_{S/C}$.

The errors from the estimation process should be checked for heteroskedasticity and autocorrelation. Logistic functions inherently produce heteroskedastic errors.[1] However, this may be counteracted by other sources of heteroskedasticity in the opposite direction. The errors should be graphed to determine the degree and direction of heteroskedasticity. Since the model has lagged dependent variables and has cross sectional data, the Durbin-Watson statistic printed by SAS is not valid. Reference [B-8] describes the test statistic applicable to this case.

Reestimation

Once the appropriate assumptions, constraints, and error structure are identified, the model should be reestimated reflecting this knowledge. If heteroskedasticity is present, the problem should be corrected, as in the previous paper, by weighted estimation.

---

1. Reference [B-1] shows the exact form of the heteroskedasticity and the proper correction.

The correction for serial correlation is more difficult. The method is described on pages 192 and 62 of [B-7]. Missing value observations should be inserted between sponsor/claimant time-series groups to avoid lagged values crossing over different sponsor/claimants.

REFERENCES

[B-1] Michael Intriligator. *Economic Models Techniques and Applications*. Englewood Cliffs, NJ: Prentice Hall, Inc., 1978

[B-2] Price Waterhouse, Contract No. N00600-86-D-3869, Delivery Order 2, *Development of an Analytical Model Relating MRRP Resources to Facility Condition Readiness*, Aug 1987

[B-3] CNA Research Memorandum 86-169, *Issues in the Measurement and Modeling of Shore Base Facility Readiness*, by Henry L. Eskew, Aug 1986 (27860169)[1]

[B-4] Dennis Aigner, Cheng Hsiao, Arie Kapteyn and Tom Wansbeek. "Latent Variable Models in Econometrics," in *Handbook of Econometrics*, edited by Zvi Griliches and Michael Intriligator. New York: Elsevier Science Publishing Co., Inc., 1984

[B-5] Daniel McFadden. "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press, 1974

[B-6] T. Amemiya, "Qualitative Response Models," *Annals of Economic and Social Measurement*, 4, 1985, pp. 363-72

[B-7] SAS Institute Inc. *SAS/ETS User's Guide*. Cary, NC: SAS Institute Inc., 1984

[B-8] J. Durbin, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressions are Lagged Dependent Variables," *Econometrica*, 38, 1970, pp. 410-21

---

1. The number in parentheses is an internal CNA control number.

(

)

APPENDIX C

TESTS FOR INTEGRATING NEW AND OLD BASEREP DATA

)

◢

APPENDIX C

TESTS FOR INTEGRATING NEW AND OLD BASEREP DATA

In the BASEREP system, shore facility readiness is reported annually for most of the Navy's shore bases. Reporting is at the activity level, and within each activity ratings are given in up to 23 mission categories.[1] In 1986, there were 1,542 ratings in the Price Waterhouse data set. The ratings consist of four readiness levels--C1, C2, C3, and C4--that indicate whether the facilities have fully, substantially, marginally, or not met the demands placed on them.

From 1982 until 1986, the activity commander, or his representative, would assign readiness ratings subjectively. A new reporting system was adopted for the 1987-ratings. There are a number of changes in the instructions for the new BASEREP system. the most important of which are revisions that make the ratings more objective. Under the new system, the person doing the rating fills out a worksheet for each mission. The worksheet gives a number of criteria and assigns a level of 1 to 4 based on how well that criteria has been met. An overall rating is assigned by choosing the lowest level, if that level appears more than once, or the next-to-lowest level, if the lowest level appears only once.

For example, for aircraft operations, there are eight criteria. One asks for what percentage of days during the year were required air operations curtailed because of the condition of runways, taxiways, arresting gear, or parking areas. A level of one is given if the percentage is no greater than 10; a level of four is given if the percentage exceeds 40.

BASEREP data for a substantial number of bases have been in existence only since 1983. It is useful to have observations over several time periods for each sponsor/claimant to estimate the relationship between funding and readiness. It is therefore tempting to combine the old and new BASEREP ratings into one time series. The CNA study team believes, however, that combining the two data series should be done cautiously if at all. There are enough differences in the way the ratings are assigned that one cannot assume that the two measures of readiness are equivalent. Some statistical tests are described below that will reveal gross discrepancies between the two measures. Even if these tests show no statistically significant differences between the old and new ratings, the two data sets should still be combined only with caution.

---

1. The number of mission categories increases from 23 to 28 under the new BASEREP instruction.

Two broad types of statistical tests for the consistency of old and new ratings are available. The first type of test would make use of the model relating funding to readiness to see whether model parameters estimated using the old data are still appropriate given the new data. The usefulness of this type of test is limited by the amount of data available and the imprecision of the facility condition readiness model. The second type of test is more general and probably more useful in this application. Tests of this type involve only the readiness data and test whether the old and new data samples are likely to have been drawn from the same underlying population.

TESTS USING THE FACILITY CONDITION READINESS MODEL

The simplest test would be to reestimate the model using the old data set plus the 1987 data. A dummy variable could be included that is set equal to zero for all observations except those with 1987 data. If the coefficient on this shift variable is significantly different from zero, then there is a shift in the relationship between funding and readiness. A negative coefficient, for example, would be evidence that with funding held constant, measured readiness is lower in 1987 than in previous years.

The problem with this test is that it tests for only one specific change in the measure of readiness. That is, it tests whether there is a constant shift in measured readiness over all sponsor/claimants that is independent of the level of funding. The presence of this shift would not be evidence that there are no other changes in the readiness measure. Thus, one would not be justified in including the dummy variable and assuming that the problem had been solved. On the other hand, if there is no such shift, one is again unable to conclude that there has been no change in measured readiness.

A more general version of this test would allow all the parameters of the model to vary between the 1983 through 1986 period and the 1987 period. This test could not be performed for a model that included sponsor/claimant intercepts, however, because there would be insufficient degrees of freedom.

Another possibility would be to perform out-of-sample tests of the model similar to those done in section II.C.3 of the Price Waterhouse report. For such tests, the model estimated using 1983 through 1986 data would be used to predict 1987 readiness given 1987 funding levels. The resulting predictions would then be compared to actual 1987 readiness. Statistics such as the root mean squared error (RMSE) or the mean absolute error (MAE) could then be calculated. If the prediction errors were large, then it could be concluded that the new data are not consistent with the model estimated using the old data.

The lack of precision in the readiness model makes this approach difficult. The section on Out-of-Sample Results in this paper points

out that the Price Waterhouse model does not perform well on out-of-sample tests even within the 1983 through 1986 period. The model does not make accurate predictions of movements in the old readiness measure. It is not clear how much worse its predictions of the new readiness measure would have to be to conclude that the two measures are inconsistent. The RMSE for 1987 could be compared to the RMSEs for 1983 through 1986. If the 1987 error is much larger, then there would be reason to believe that the 1987 data should not be combined with the earlier data.

The advantage of tests using the facilities condition readiness model is that they control for changes in readiness explained by variables in the model. Unfortunately, for a sponsor/claimant model, there is not enough new data to make general tests possible for structural shifts in the model's parameters. Also, tests that compare predicted to actual readiness are problematical because of the low predictive power of the model.

TESTS USING THE READINESS DATA

More general tests would compare only the new and old readiness data, without reference to the readiness model. These tests are more general and perhaps more reliable given the quality of the readiness model. On the other hand, changes in actual readiness caused by changes in funding cannot be distinguished from changes in measured readiness. Both parametric and nonparametric tests are described in this section. Also, the type of test that is used would depend on the level of aggregation of the data.

A Nonparametric Test Using Disaggregated Data

At the lowest level of aggregation, each mission within an activity has a readiness measure of C1, C2, C3, or C4. Since this is an ordinal measure, it does not make sense to compute mean levels of readiness across all ratings and compare them between 1986 and 1987. There is, however, a nonparametric test of whether ratings tend to remain about the same between 1986 and 1987. This test is referred to as the sign test.[1]

For the sign test, the readiness ratings in 1986 and 1987 are compared for each individual mission in each activity. If the 1987 rating is higher, a positive sign is assigned; if the 1987 rating is lower, a negative sign is assigned. If the ratings are the same, the pair is discarded and the sample size is reduced. A test statistic is computed by counting the number of positive signs, S, and the number of missions in which readiness did not remain constant, n. The test statistic for the null hypothesis that 1987 ratings are just as likely to be below as

---

1. See Gouri K. Bhattacharyya and Richard A. Johnson, *Statistical Concepts and Methods* (New York: John Wiley & Sons, 1977), pp. 516-19.

above 1986 ratings, or that the probability of a positive sign is 0.5, is given by

$$Z = \frac{S - n/2}{\sqrt{n/4}} \; . \tag{C-1}$$

For large samples (for example, n above 30), this test statistic has a standard normal distribution under the null hypothesis. A two-tailed test should be performed, since there is no reason to believe that the difference between the new and old ratings would be either positive or negative.

The test can be improved by comparing the test statistic for 1986 versus 1987 to similar statistics for 1985 versus 1986 and for other pairs of years. The test statistics for the earlier years would give base values for how much change to expect in the readiness ratings from year to year. If the 1986 versus 1987 test statistics are much larger than the others, then there would be reason to believe that the 1987 data are different. A qualification is that there may have been a change in actual readiness levels between 1986 and 1987. The sign test cannot distinguish between changes in actual readiness and changes in the measure of readiness.

Tests on Aggregated Data

If the readiness data are aggregated above the individual mission level, then other tests can be used. One measure that is frequently used is the percentage of ratings that are either C1 or C2. This percentage can be calculated over all the ratings for an activity, or sponsor/claimant, or the entire Navy. In general, there will be more information contained in statistics calcula'd with less aggregated data. With the readiness ratings translated into a cardinal measure, however, a broader range of tests is available.

A nonparametric test that is similar to the sign test but also takes into account the relative magnitude of differences between 1986 and 1987 ratings is the Wilcoxon signed-rank test.[1] Parametric tests would include calculating correlation coefficients between 1986 and 1987 readiness measures and testing the equality of sample means and variances. If these tests show no significant difference between measured readiness in 1986 and 1987, it lends some support to combining the data series. There is no test, however, that can prove conclusively that the new and old readiness measures are equivalent.

---

1. Ibid., pp. 512-23.